

## 1 Multinomial Logit (MNL) model

Special case of MNL model for the binary response variable is Logistic regression.

If  $U_i = V_i + \epsilon_i, \forall i \in \mathcal{N}$ , where  $V_i$  is the deterministic component of the utility and  $(\epsilon_i)_{i=1}^n$  are i.i.d.  $Gumbel(0, \mu)$ , the choice probability is given by

$$\mathbb{P}(i|\mathcal{S}) = \frac{e^{\mu V_i}}{\sum_{j \in \mathcal{S}} e^{\mu V_j}} \text{ MNL choice probability.} \quad (1)$$

If we set  $v_i = e^{\mu V_i}$ , then we get  $\mathbb{P} = \frac{v_i}{\sum_{j \in \mathcal{S}} v_j}$ .

## 2 Attraction model

The attraction model is a slight generalization of the MNL model. Here, it is assumed that every product  $j \in \mathcal{N}$  has a non-negative “attraction” denoted by  $v_j$ . Now, the probability that item  $i \in \mathcal{N}$  will be chosen from offer set  $\mathcal{S}$  is proportional to its attractiveness.

$$\mathbb{P}(i|\mathcal{S}) \propto v_i \Rightarrow \mathbb{P}(i|\mathcal{S}) = \frac{v_i}{\sum_{j \in \mathcal{S}} v_j}. \quad (2)$$

For example, suppose  $\mathcal{N} = \{1, 2, 3\}$  where 1: Bus, 2: Train, 3: Car. We have two types of people: Type 1:  $\{1, 2\}$ , Type 2:  $\{2, 3\}$ .

Note: The attraction model allows the attraction parameters  $v_j$  to be equal to zero. This is unlike the MNL model, where all the parameters  $v_j = e^{\mu V_j} > 0$ . Therefore, the attraction model generalizes the NML model. This model is widely used in revenue management.

For example,  $\mathcal{N} = \{0, 1, 2\}$  where 0 stands for “Not purchase option”.

$$\mathbb{P}(1|\{0, 1, 2\}) = \frac{v_1}{v_0 + v_1 + v_2}, \quad (3)$$

$$Pr(purchase) = \frac{v_1 + v_2}{v_0 + v_1 + v_2}. \quad (4)$$

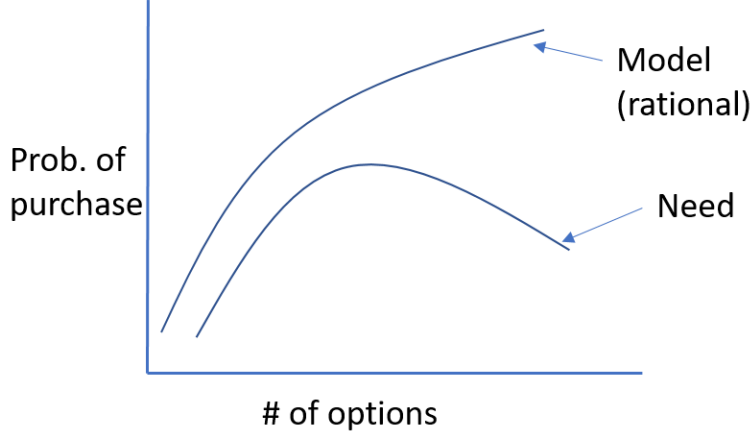


Figure 1: Purchase curve

### 3 Plackett-Luce model

R. Duncan Luce is a renowned mathematical psychologist. In 1959 (before RUM framework was formalized and MNL model existed), Luce adopted the axiomatic approach to choice. He started with the stochastic choice function  $\mathbb{P}(i|\mathcal{S})$  for  $i \in \mathcal{S}$  and  $\mathcal{S} \in \mathcal{N}$ , and imposed the following conditions:

- (a) for any  $i \in \mathcal{S}$  and  $\mathcal{S} \in \mathcal{N}$ ,  $0 \leq \mathbb{P}(i|\mathcal{S}) \leq 1$ ,
- (b)  $\mathbb{P}(\mathcal{S}|\mathcal{S}) = 1$  where  $\mathbb{P}(\mathcal{M}|\mathcal{S}) = \sum_{j \in \mathcal{M}} \mathbb{P}(j|\mathcal{S})$ ,
- (c) If  $\mathcal{C}_1, \mathcal{C}_2$  s.t.  $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$  and  $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathcal{S}$ , then  $\mathbb{P}(\mathcal{C}_1 \cup \mathcal{C}_2|\mathcal{S}) = \mathbb{P}(\mathcal{C}_1|\mathcal{S}) + \mathbb{P}(\mathcal{C}_2|\mathcal{S})$ ,
- (d) [Luce's Choice Axiom]:
  - (i) if  $\mathbb{P}(i|\{i, j\}) \neq 0 \forall i, j \in \mathcal{N}$ , then  $\forall \mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \mathcal{S}$ , we must have that

$$\mathbb{P}(\mathcal{C}_1|\mathcal{S}) = \mathbb{P}(\mathcal{C}_2|\mathcal{S}) \cdot \mathbb{P}(\mathcal{C}_1|\mathcal{C}_2), \quad (5)$$

- (ii) if  $\mathbb{P}(i|\{i, j\}) = 0$  for some  $i, j \in \mathcal{S}$ , then for any  $\mathcal{C} \subseteq \mathcal{S}$ , we must have

$$\mathbb{P}(\mathcal{C}|\mathcal{S}) = \mathbb{P}(\mathcal{C} \setminus \{i\}|\mathcal{S} \setminus \{i\}). \quad (6)$$

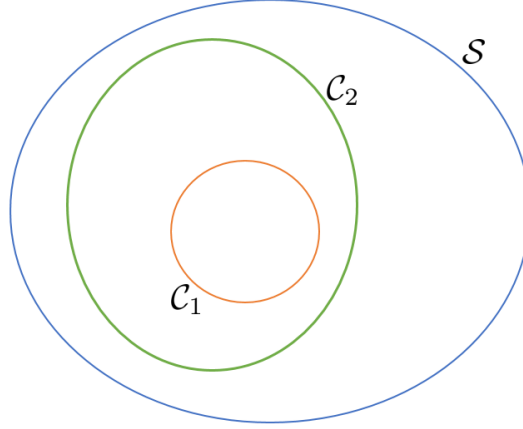


Figure 2: Example of  $\mathcal{S}, \mathcal{C}_1, \mathcal{C}_2$

Let  $E_1$ : choice is in set  $\mathcal{C}_1$  when  $\mathcal{S}$  is offered,  $E_2$ : choice is in set  $\mathcal{C}_2$  when  $\mathcal{S}$  is offered,  $E_1 \subseteq E_2$ , then

$$\mathbb{P}(\mathcal{C}_1|\mathcal{S}) = Pr(E_1) \quad (7)$$

$$= Pr(E_1 \cap E_2) \quad (8)$$

$$= Pr(E_1|E_2) \cdot Pr(E_2) \quad (9)$$

$$= Pr(E_1|E_2) \cdot Pr(\mathcal{C}_2|\mathcal{S}). \quad (10)$$

Let  $E$ : choice is in set  $\mathcal{C}_1$  when offered  $\mathcal{C}_2$ , then

$$Pr(E) = \mathbb{P}(\mathcal{C}_1|\mathcal{C}_2). \quad (11)$$

In other words, Luce's Axiom is imposing that

$$Pr(E_1|E_2) = Pr(E). \quad (12)$$

Writing (12) in English is that  $Pr(\text{choice in } \mathcal{C}_1 \text{ when } \mathcal{S} \text{ offered} \mid \text{choice in } \mathcal{C}_2 \text{ when } \mathcal{S} \text{ offered}) = Pr(\text{choice in } \mathcal{C}_1 \text{ when } \mathcal{C}_2 \text{ is offered})$ , which means that rest of items are irrelevant if knowing choice is from  $\mathcal{C}_2$ .

For example,  $\mathcal{S} = \{\text{car, bike, train, bus}\}$ ,  $\mathcal{C}_1 = \{\text{bus}\}$ ,  $\mathcal{C}_2 = \{\text{bus, train}\}$ .

Conditioned on the fact that the choice belongs to  $\mathcal{C}_2$ , the probability that the choice belongs to  $\mathcal{C}_1$  does not depend on the larger set  $\mathcal{S}$ . Therefore, for the purpose of choosing from the set  $\mathcal{C}_2$ , the elements in the set  $\mathcal{S} \setminus \mathcal{C}_2$  become irrelevant. Because of this, this property is also called the **Independent of Irrelevant Alternatives** (IIA) property.

IIA is a term that was used by Kenneth Arrow in his set of axioms. It was explained through the following example:

Initial set: {apple pie, blueberry pie}

Final set: {apple pie, blueberry pie, cherry pie}

**Proposition 1** For any two items,  $i, j \in \mathcal{S}_1, \mathcal{S}_2$ , then the choice axiom implies that

$$\frac{\mathbb{P}(i|\mathcal{S}_1)}{\mathbb{P}(j|\mathcal{S}_1)} = \frac{\mathbb{P}(i|\mathcal{S}_2)}{\mathbb{P}(j|\mathcal{S}_2)}. \quad (13)$$

*Proof:* Let  $\mathcal{C}_1 = \{i\}$ ,  $\mathcal{C}_2 = \{i, j\}$ ,  $\mathcal{S} = \mathcal{S}_1$ . By Luce's choice axiom, we have

$$\mathbb{P}(i|\mathcal{S}_1) = \mathbb{P}(\mathcal{C}_2|\mathcal{S}) \cdot \mathbb{P}(\mathcal{C}_1|\mathcal{C}_2) = \mathbb{P}(\{i, j\}|\mathcal{S}_1) \cdot \mathbb{P}(i|\{i, j\}). \quad (14)$$

Similarly, we get

$$\mathbb{P}(j|\mathcal{S}_1) = \mathbb{P}(\{i, j\}|\mathcal{S}_1) \cdot \mathbb{P}(j|\{i, j\}). \quad (15)$$

$$\Rightarrow \frac{\mathbb{P}(i|\mathcal{S}_1)}{\mathbb{P}(j|\mathcal{S}_1)} = \frac{\mathbb{P}(\{i, j\}|\mathcal{S}_1) \cdot \mathbb{P}(i|\{i, j\})}{\mathbb{P}(\{i, j\}|\mathcal{S}_1) \cdot \mathbb{P}(j|\{i, j\})} = \frac{\mathbb{P}(i|\{i, j\})}{\mathbb{P}(j|\{i, j\})}. \quad (16)$$

In a similar fashion, we can show that

$$\frac{\mathbb{P}(i|\mathcal{S}_2)}{\mathbb{P}(j|\mathcal{S}_2)} = \frac{\mathbb{P}(i|\{i, j\})}{\mathbb{P}(j|\{i, j\})}. \quad (17)$$

□

It can be shown that Luce's choice model is equivalent to the MNL model. Hint: (Go from Luce's to MNL) Set  $v_i = \mathbb{P}(i|\mathcal{N})$  & choice axiom  $\Rightarrow \mathbb{P}(i|\mathcal{S}) = \frac{v_i}{\sum_{j \in \mathcal{S}} v_j}$ .

## 4 Red bus - Blue bus example

Setting:

- Bus, Train: People are indifferent between the two. 50% take Bus and 50% take Train.
- Bus: Split into two choices: red bus & blue bus. If we offer red and blue buses, then customer split evenly between the two.

Now suppose we offer all the three options: {Red bus, Blue bus, Train}. 25% choose red bus, 25% choose blue bus, and the remaining 50% choose the train.

$\mathcal{S} = \{R, B, T\}$ ,  $\mathcal{C}_2 = \{B, T\}$ ,  $\mathcal{C}_1 = \{B\}$

Given the choice probability from offering  $\mathcal{S}$ ,  $Pr(\mathcal{C}_1|\mathcal{S}) = \frac{1}{4}$ ,  $Pr(\mathcal{C}_2|\mathcal{S}) = \frac{3}{4}$ . By the choice axiom, it follows that

$$\mathbb{P}(\mathcal{C}_1|\mathcal{C}_2) = \frac{\mathbb{P}(\mathcal{C}_1|\mathcal{S})}{\mathbb{P}(\mathcal{C}_2|\mathcal{S})} = \frac{1/4}{3/4} = \frac{1}{3}. \quad (18)$$

$\Rightarrow Pr(\text{Blue bus is chosen from Blue bus \& Train}) = \frac{1}{3}$ .

An implication of Luce's Axiom is proportional substitution, i.e., when an option is dropped from the offer set, customer who used to buy that option substitute to other items in proportion to their corresponding market shares.

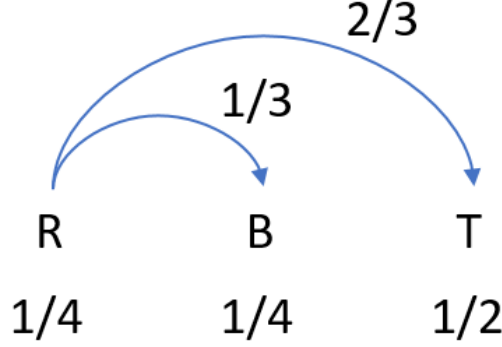


Figure 3: Substitution under Luce's Axiom

## 5 Taking the MNL model to practice

So far, we have focused on  $\epsilon$  (the error term). We now focus on the deterministic component,  $V$ .

One of key functions of the MNL model is to understand how various factors affect choices of customers. It can be used to (a) explain choice in terms of factors affecting it; (b) run counter factials: what would happend if price or another feature is changed; (c) allows one to put a \$ value on different attributes.

Let's assume the following:

$$utility = f(attributes, price). \quad (19)$$

We now express the deterministic component of the utility in terms of observable features. We distinguish two types of features: (a) alternative-specific: price, color, packaging, flavors; (b) individual/choice instance specific: income, gender, time of day, geographic location, etc.

We typically express this as

$$V_{ij} = f(\underline{z}_{ij}, \underline{s}_i), \quad (20)$$

where

- $V_{ij}$ : Deterministic utility for item  $j$  in choice instance  $i$ .
- $\underline{z}_{ij}$ : Vector of alternative specific attributes.
- $\underline{s}_i$ : Vector of individual/choice instance specific attributes.

We use the standard assumption that  $f(\cdot, \cdot)$  is linear-in-parameters. In other words,  $f(\cdot, \cdot)$  must be linear in the parameters that describe the model and not necessarily linear in attributes.

$$V_{ij} = \underline{\beta}^\top \begin{bmatrix} \underline{z}_{ij} \\ \underline{s}_i \end{bmatrix}, \quad (21)$$

where  $\underline{\beta}$  are parameters and  $\begin{bmatrix} \underline{z}_{ij} \\ \underline{s}_i \end{bmatrix}$  are attributes.

We distinguish two types of coefficients:

- (a) generic coefficients: coefficients that stay the same across the alternatives.
- (b) alternative specific coefficients: coefficients specific to each alternative.

Therefore, we can write (21) into the following form,

$$V_{ij} = \alpha_j + \underline{\beta}^\top \underline{z}_{ij} + \underline{\gamma}_j^\top \underline{s}_i + \underline{\delta}_j^\top \underline{w}_{ij}, \quad (22)$$

where

- $\alpha_j$ : intercept term/ Alternative Specific Constant (ASC).
- $\underline{\beta}$ : generic coefficients (does not depend on  $j$ ).
- $\underline{\beta}^\top \underline{z}_{ij}$ : generic coefficients - alternative specific attributes ( $\beta \times Price_{ij}$ ).
- $\underline{\gamma}_j$ : alternative specific coefficients (depend on  $j$ ).
- $\underline{\gamma}_j^\top \underline{s}_i$ : alternative specific coefficients - individual specific attributes ( $\gamma_j \times Income_i$ ).
- $\underline{\delta}_j$ : alternative specific coefficients.
- $\underline{\delta}_j^\top \underline{w}_{ij}$ : alternative specific coefficients - alternative specific attributes ( $\delta_j \times Price_{ij}$ ).